AD 740830

DEPARTMENT OF STATISTICS

HARVARD UNIVERSITY

AN EMPIRICAL STUDY OF THE HALF-NORMAL PLOT

by

DOUGLAS A. ZAHN

Technical Report No. 35

September 1, 1971

Contract N00014-67A-0298-0017, NR-042-097

This paper is the result of research supported by the Office of Naval Research. Reproduction in whole or in part is permitted by the U.S. Government. Distribution of this document is unlimited.

45

Unclassified

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Harvard University Department of Statistics Cambridge, Massachusetts	2a. REPORT SECURITY CLASSIFICATION
	2b. GROUP

3. REPORT TITLE

AN EMPIRICAL STUDY OF THE HALF-NORMAL PLOT

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Report; September 1, 1971

5. AUTHOR(S) (Last name, first name, initial)

ZAHN DOUGLAS A.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
September 1, 1971	43	14

8a. CONTRACT OR GRANT NO. N00014-67A-0298-0017	9a. ORIGINATOR'S REPORT NUMBER(S) Technical Report No. 35
b. PROJECT NO. NR-042-097	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Logistics & Mathematical Statistics Branch Office of Naval Research Dept. of the Navy, Washington, D.C.

13. ABSTRACT

Monte Carlo studies of the original version of the half-normal plot (Daniel, Technometrics 1 (1959) 311-341) and two new versions are reported. Data representative of the 15 contrasts from a 2^{p-q}, p-q = 4, factorial experiment are generated. Design parameters are α, the experimentwise error rate, r, the number of real contrasts present in the 2^4 experiment, and m, the size of the real contrasts present. Studies are made for designs with α = .05, .20, .40, r = 0, 1, 2, 4, and 6, m = $0(2\sigma)8\sigma$, where σ is the standard deviation of a contrast.

We give critical values for the various versions which control the experimentwise error rate. These critical values are considerably different than those given by Daniel.

Detection rate, i.e., the proportion of real contrasts declared significant, false positive behavior, and estimation of σ are examined. The Monte Carlo studies indicate that one of the new versions is superior to the original version. The detection rate of all versions decreases drastically when r increases from one to two to four. When several small real contrasts are present, the sensitivity can be increased and the magnitude of the average errors in estimating σ can be greatly reduced by using α = .2 or .4, rather than α = .05.

Nomination procedures for analyzing a single replicate 2^4 factorial experiments have a smaller detection rate than the half-normal plot with an equivalent experimentwise error rate, unless the experimenter can accurately nominate ten error contrasts in the 2^4 experiment.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

AN EMPIRICAL STUDY OF THE HALF-NORMAL PLOT

by

Douglas A. Zahn
Florida State University

1. INTRODUCTION

The half-normal plot as introduced by Daniel (1959) is a multi-purpose tool for criticizing and interpreting the contrasts yielded by a single replication factorial experiment. It gives us a means for approaching the following general problem:

> Let the statistics $Y_1, Y_2, \ldots, Y_n$ be independent, normally distributed random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and common variance σ^2. Assume that at most r of the μ_i are non-zero, where r is small relative to n, and that we have no prior estimate of σ. Given $y_1, y_2, \ldots, y_n$ (single observations of each of the n random variables), (1) decide which, if any, of the μ_i are non-zero and (2) estimate σ.

This problem arises in many contexts. The motivating one was single replication factorial experiments where we can think of y_i as a contrast estimating the effect μ_i of a factor or interaction between factors in the experiment. The half-normal plot can be used to decide which of the μ_i are significantly non-zero. After the significant contrasts have been isolated,we can then use the half-normal plot to estimate σ from the insignificant contrasts.

Birnbaum (1959) investigated the probability that the half-normal

plot will detect a single real contrast, i.e., a contrast with mean $\mu \neq 0$, when $n = 31$. He found that it compared well to the multiple t-test in this situation. However, he warned that his findings rested heavily on the assumption that only one real contrast was present. He also warned that if more than one real contrast was present, the power of the plot may be greatly reduced. This report presents results of a Monte Carlo study of Daniel's original version of the half-normal plot and two modifications of it when as many as six real contrasts of size 1σ to 8σ are present in a 2^{p-q}, $p-q = 4$, factorial experiment. Thus, we consider the power, false positive behavior, and variance estimation of these versions of the half-normal plot when they are applied to the general problem in the case $n = 15$, since there are 15 contrasts of interest, ignoring the grand mean, in a 2^4 factorial experiment. However these versions, or obvious modifications of them, can be used for any n. The cases $n = 8, 9, \ldots, 15$ can be analyzed using critical values in Table 3.1. Critical values for other cases can be constructed using procedures described in Section 3. Additional results are presented in Zahn (1969) and are available from the author on request.

2. VERSIONS OF THE HALF-NORMAL PLOT

To investigate the effects of modifying various steps of the half-normal plot on sensitivity and variance estimates, we have included a comparison of several versions of the half-normal plot in our Monte Carlo investigation.

2.1 *Version* X

Calculate the order statistics, $x_1 \leq x_2 \leq \cdots \leq x_{15}$, of the absolute values of the 15 observed contrasts. Divide the order statistics by s_i, the initial estimate of σ. In this version $s_i = x_{11}$. This produces a set of standardized order statistics, $t_1 \leq t_2 \leq \cdots \leq t_{15}$, which are also used as test statistics to examine the order statistics for significance.

This set of scale-free order statistics may be plotted on a revised standardized half-normal grid, illustrated in Figure 2.1. The .05, .20, and .40 level critical values for this version are indicated by the points on the grid which are connected by lines to form the .05, .20, and .40 level guardrails, respectively. This grid differs from the standardized half-normal grid given in Daniel (1959) in three ways:

(1) The standardized order statistics are plotted as the ordinate, rather than as the abscissa, as in Daniel (1959), to make the half-normal plot correspond more closely to the usual regression graph on which the random variable is plotted as the ordinate.

(2) To take advantage of benefits cited by Ferguson (1960), the standardized order statistics are plotted against the mean of the order statistics of a random sample of size 15 from the standard half-normal distribution. These means have been computed by Blankenship (1965) and are given in Table 2.1 for samples of sizes 1(1)15.

(3) The guardrails given on the revised grid differ considerably from those on the grids in Daniel (1959).

FIGURE 2.1

Revised Standardized Half-Normal Grid for Version X for 15 Contrasts

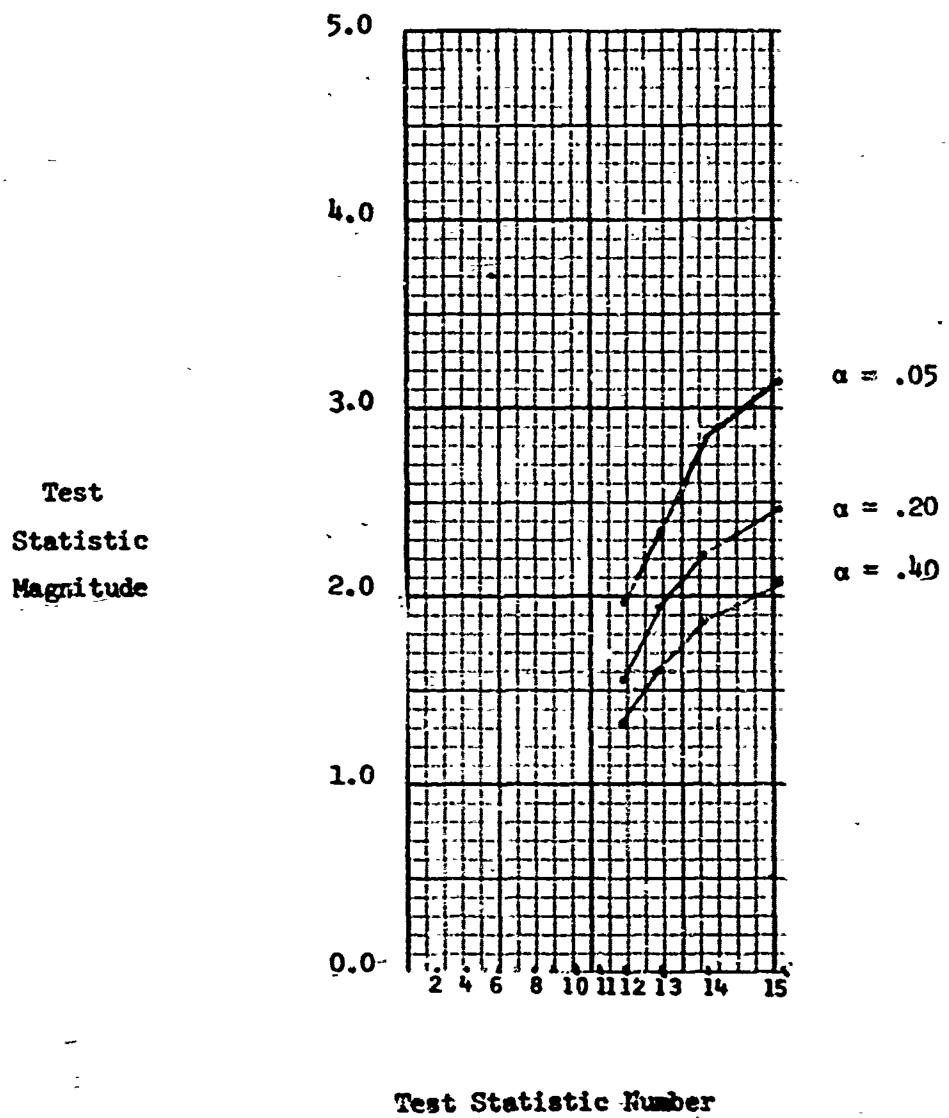


TABLE 2.1

Means of the Order Statistics of Random Samples of Sizes 1, 2, ..., 7 From the Standard Half-Normal Distribution

The table entry in row i and column j is m(i,j) = the mean of the i^{th} order statistic in a random sample of size j from the standard half-normal distribution.

i \ j	1	2	3	4	5	6	7
1	0.798	0.467	0.335	0.262	0.216	0.183	0.160
2		1.128	0.732	0.553	0.448	0.377	0.326
3			1.326	0.911	0.712	0.589	0.504
4				1.465	1.044	0.835	0.702
5					1.570	1.149	0.934
6						1.654	1.235
7							1.724

TABLE 2.1, continued

Means of the Order Statistics of Random Samples of Sizes 8, 9, ..., 15 From the Standard Half-Normal Distribution

The table entry in row i and column j is m(i,j) = the mean of the i^{th} order statistic in a random sample of size j from the standard half-normal distribution.

i \ j	8	9	10	11	12	13	14	15
1	0.14	0.13	0.12	0.11	0.10	0.09	0.08	0.08
2	0.29	0.26	0.23	0.21	0.20	0.18	0.17	0.16
3	0.44	0.39	0.35	0.32	0.30	0.27	0.26	0.24
4	0.61	0.54	0.48	0.44	0.40	0.37	0.34	0.32
5	0.80	0.70	0.62	0.56	0.51	0.47	0.44	0.41
6	1.02	0.88	0.77	0.69	0.63	0.58	0.53	0.50
7	1.31	1.09	0.94	0.84	0.76	0.69	0.64	0.59
8	1.78	1.37	1.15	1.00	0.90	0.81	0.74	0.69
9		1.84	1.42	1.21	1.06	0.95	0.86	0.79
10			1.88	1.47	1.25	1.11	1.00	0.91
11				1.92	1.52	1.30	1.15	1.04
12					1.96	1.56	1.34	1.19
13						1.99	1.59	1.38
14							2.02	1.63
15								2.05

The computation of our guardrails and the error rate they are intended to control are considered in Section 3. Table 3.1 gives the critical values on which the guardrails for this version and the subsequent two versions are based.

The detection process is a sequential statistical procedure in that whether contrasts are tested for significance late in the process depends on results observed early in the process. We first test x_{15} for significance by comparing test statistic 15 = t_{15} = x_{15}/x_{11} to the appropriate critical value c_{15}. If $t_{15} > c_{15}$, x_{15} is declared significant and we then examine t_{14}. We continue testing contrasts until one is declared insignificant or until a maximum of four contrasts are declared significant.

If desired, the detection process can be carried out without graphing the order statistics. We need merely compute the appropriate test statistics and compare them to the critical values. Of course, the beauty of the plot is that it also enables us to examine the contrasts for the abnormalities discussed by Daniel (1959).

The detection process divides the contrasts into two sets: significant contrasts and insignificant contrasts. The latter set will be referred to as error contrasts because we calculate s_f, our final estimate of σ, from them. Let e denote the number of error contrasts. We plot on ordinary linear-by-linear graph paper $x_1, x_2, \ldots, x_e$ against $m(i,e)$, $i = 1, 2, \ldots, e$, where $m(i,e)$ denotes the mean of the i^{th} order statistic in a random sample of size e from the standard half-normal distribution. We fit the least squares regression line through the origin of x on m. The slope of this line

is s_f, the final estimate of σ, which should not be confused with s_i, the initial estimate of σ.

The cases n = 14, 13, and 12 can also be examined using this version with x_{11} as the test statistic denominator and the critical values in Table 3.1, provided that the researcher is willing to assume that no more than three, two, or one real contrasts are present, respectively. For instance, if n = 13, the first test statistic is $t_{13} = x_{13}/x_{11}$ and its .05 level critical value is 2.36; the second is $t_{12} = x_{12}/x_{11}$ and its .05 level critical value is 1.94.

2.2 *Version* S

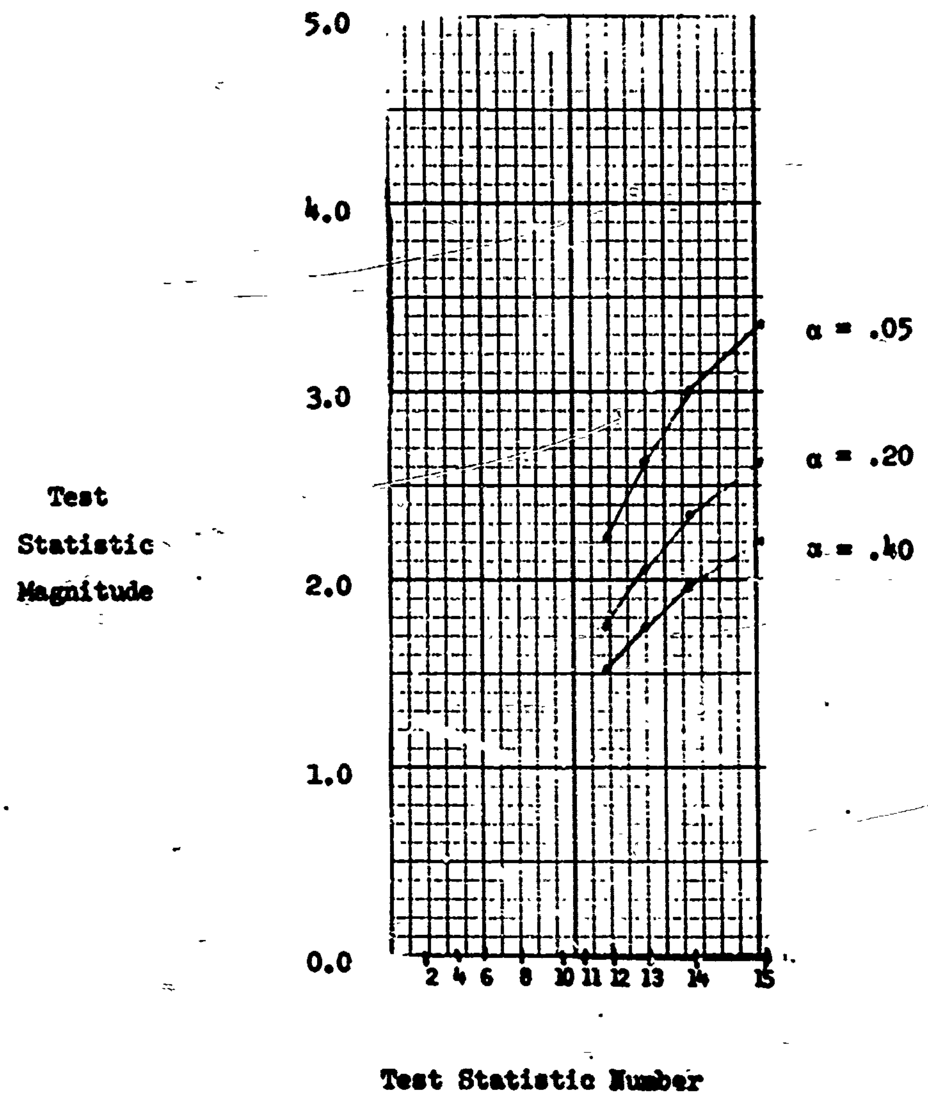
Versions S and X only differ with respect to the test-statistic denominators used during the detection process. We define

$$SL(k,j) = \sum_{i=1}^{k} x_i m(i,j) \Big/ \sum_{i=1}^{k} [m(i,j)]^2, \; k \leq j.$$

The statistic SL(11,15) is used as test statistic denominator by version S. This is the slope of the least squares regression line through the origin of x on m fitted to the points $[m(i,15),x_i]$, i = 1, 2, ..., 11. For this version, the standardized order statistics are $t_{15} = x_{15}/SL(11,15)$, etc. Thus, the test statistic denominators of version S should be less variable estimators of σ than the test statistic denominators of X, since the denominators of S are based on more information. Hence, the guardrails on the revised standardized half-normal grid for version S illustrated in Figure 2.2 differ from those on the grid for version X. Again, four is the maximum number of contrasts which may be declared significant. The final estimate

FIGURE 2.2

Revised Standardized Half-Normal Grid for Version 8 for 15 Contrasts



of σ is obtained exactly as it is by version X. Using the above notation we see that, if there are e error contrasts, $s_r = SL(e,e)$.

2.3 *Version* R

Only one of three other half-normal plot versions investigated in our Monte Carlo study will be considered here. The first step in version R is to compute the standardized order statistics $t_i = x_i/SL(7,15)$, $i = 1, 2, \ldots, 15$. Here the test statistic denominator is the least squares line through the origin fitted to the smallest half of the contrasts being tested for significance. If x_{15} is declared significant, the remainder of the detection process in this version differs considerably from the detection process in versions X and S.

We reassess our position every time a contrast is declared significant. This step reflects the fact that a contrast appearing significantly large alters our assessment of the state of nature. If x_{15} is declared significant, we then examine the remaining 14 absolute contrasts under the hypothesis that they constitute a random sample of size 14 from the half-normal distribution. These 14 order statistics are restandardized by dividing by $SL(7,14)$ which is an unbiased estimate of σ under this hypothesis. We now compare the restandardized value of x_{14} to the appropriate critical value. If it is declared significant, we concentrate on the remaining 13 absolute contrasts, considering them as a random sample of size 13 from the half-normal distribution. We restandardize them by dividing by $SL(6,13)$. If x_{14} is not declared significant, s_r is computed as before from the 14 error contrasts.

Another difference between this version and the previous two is that R may declare as many as seven contrasts significant. To gain this additional flexibility, we have based on the test statistic denominators on at most the seven smallest absolute contrasts.

3. CRITICAL VALUES FOR THE HALF-NORMAL PLOT VERSIONS

Before any version can be used, we need critical values for judging the test statistics. Clearly, each version requires several critical values since each may declare more than one contrast significant. The α level critical value for a given test statistic t is defined to be the $(1-\alpha)$ quantile of its distribution under the hypothesis H: all contrasts currently being tested for significance have means $\mu = 0$. For example, for $t_{13} = x_{13}/x_{11}$ for version X, the .05 level critical value is the .95 quantile of its distribution under the hypothesis H: the 13 contrasts being tested for significance all have means $\mu = 0$. Table 3.1 gives .05, .20, and .40 level critical values for all versions. We determined these critical values by generating an empirical distribution based on 999 simulations for each test statistic and estimating the .95, .80, and .60 quantiles by the corresponding quantiles of the empirical distribution.

The precision of our estimated quantiles can be evaluated using methods described by Wilks (1962, p. 331). We are 90% confident that the critical values given in Table 3.1 are correct to within $\pm$ .10, except for the .05 level critical values for version R. We are 90% confident that these values are correct to within $\pm$ .25.

TABLE 3.1

.05, .20, and .40 Level Critical Values of the Test Statistics Used By All Versions.

These α level critical values are the corresponding (1-α) quantiles of the simulated empirical distributions.

Version	Level = α	Test Statistic						
		t_{15}	t_{14}	t_{13}	t_{12}	t_{11}	t_{10}	t_9
R	.05	4.13	4.11	4.31	4.17	4.43	4.15	4.69
	.20	2.94	2.89	2.96	2.92	2.94	2.84	2.90
	.40	2.32	2.26	2.27	2.22	2.22	2.15	2.13
X	.05	3.14	2.83	2.36	1.94			
	.20	2.48	2.21	1.91	1.54			
	.40	2.08	1.87	1.60	1.31			
S	.05	3.37	3.09	2.61	2.21			
	.20	2.61	2.34	2.06	1.76			
	.40	2.20	1.97	1.76	1.51			

What aspect of the false positive behavior of a version is controlled by the critical values used obviously depends on the choice of critical values. The aspect we chose to control is the error rate per experiment (EER). The EER is defined to be the probability of declaring at least one false positive in the analysis of one experiment. It is easy to see that these critical values control the EER if no real contrasts are present. In this situation we declare at least one false positive if and only if we declare the largest contrast significant. This occurs if t_n, the test statistic for examining x_n, is larger than c_n, its α level critical value. But, c_n is the $(1-\alpha)$ quantile of t_n under the hypothesis that all n contrasts being examined have means $\mu = 0$. Hence, EER = $P(t_n > c_n) = \alpha$.

Using the $(1-\alpha)$ quantiles as α level critical values in general situations, we have proved the following theorem for $n = 2$ and $n = 3$.

> <u>Theorem</u>: The experimentwise error rate (EER) of the half-normal plot using α level critical values is $\leq \alpha$, regardless how many real contrasts of various sizes are present in the experiment being analyzed.

The proofs for these cases and much empirical support for the case $n = 15$ are present in Zahn (1969).

3.1. *Differences Between the Empirically Determined Critical Values and Daniel's Critical Values*

Since version X and the original version of the half-normal plot use identical test statistics, critical values corresponding to those estimated by the simulation results for version X can be read from the

guardrails of Figure 11a of Daniel (1959). When no real contrasts are present, Daniel's guardrails and the simulated guardrails will yield approximately the same EER. However, if one large contrast is present, Daniel's .05 guardrail may yield an EER as large as .20. Daniel refers to his .05 guardrail as having a "rejection rate" of .05. The large difference between the "rejection rate" of Daniel's guardrails and the EER actually yielded by those guardrails persists when more than one real contrast is present. An experimenter using Daniel's original version of the half-normal plot should definitely be aware of this deficiency in the critical values presented in Table 11a of Daniel (1959).

4. THE MAIN SIMULATION STUDY

Using the critical values from Section 3, we have performed computer sampling experiments to investigate the detection rate, error rates, and variance estimation of the half-normal plot and to compare the three versions described in Section 2.

4.1 *Situations Examined*

This section defines the notation used to describe a given "situation" and lists all situatons which were examined. A situation is a specification of the number and sizes of the real contrasts present in the experiment. We assume that the state of nature and the experimental design are such that all Model I Anova assumptions are satisfied.

In the simplest situation, the null situation, all $\mu_i = 0$. The null situation, denoted (0), represents the state of nature when the

null hypothesis that all contrasts are null is true. For a given situation representing a particular alternative hypothesis, one or more of the $\mu_i \neq 0$. Daniel (1959) and Birnbaum (1959) investigated situations in which one μ_i ranged from 1σ to 6σ. We consider situations in which from one to six of the μ_i are non-zero. A situation with as many as six real contrasts might occur, for instance, when an experimenter encounters a 2^{6-2} fractional factorial experiment in which four main effects, along with two of the two-factor interactions, are non-zero.

With more than one real contrast present, the situations easiest to characterize are those in which contrasts of only one size are present. We refer to these as Type I situations and consider them in Section 5. We define (r,m) to be a situation in which there are r real contrasts, each of size m. Thus $(4,6\sigma)$ implies four non-zero contrasts, each of size 6σ.

In Section 6 we consider Type II situations, i.e., situations with real contrasts of two different sizes present. We define $(r_1,m_1;r_2,m_2)$ to be a situation in which there are r_1 real contrasts, each with mean m_1, and r_2 real contrasts, each with mean m_2. For instance, $(2,6\sigma;2,8\sigma)$ implies four non-zero contrasts, two of size 6σ and two of size 8σ. The remaining eleven contrasts are null, i.e., each has mean zero.

We define an r-situation to be a Type I situation with r non-zero contrasts. Thus, $(4,6\sigma)$ is a particular 4-situation. Similarly, we define an r_1-r_2-situation to be a Type II situation in which there are r_1 contrasts of size $m_1 \neq 0.0$ and r_2 contrasts of size $m_2 \neq 0.0$.

The Main Study examined Type I situations (r,m), where $r = 1, 2, 4, 6$ and $m = 0(2\sigma)8\sigma$. We also examined Type II situations $(r_1,m_1;r_2,m_2)$, where $r_1 = r_2 = 1, 2$; $m_1 = 2\sigma, 4\sigma, 6\sigma$; and $m_1 \leq m_2 = 4\sigma, 6\sigma, 8\sigma$. We simulated each situation 1000 times and analyzed the simulated data using each version with each of three critical value levels: $\alpha = 0.05, 0.20$, and 0.40. To discuss the performance of version S in analyzing experiments with 2 real contrasts of size 6σ present, we will, for brevity, refer to results for S in $(2,6\sigma)$.

The pseudo-random standard normal deviates used in our simulation studies were generated by the Harvard Computing Center's RANDOM function subroutine which is available on the IBSYS, Fortran IV system using an IBM 7090/4 computer.

4.2 *The Experimental Design*

The Main Study may be viewed as a factorial experiment in which there are three factors, version, α, and situation, at 3, 3, and 29 levels, respectively. For each situation we decided to examine the three versions at each α using the <u>same</u> 1000 sets of 15 rsnd. Hence, in each situation we introduced a positive correlation between the results for each version and increased the precision of comparisons among the various versions in the same situation. This design is analogous to a split-plot or nested design in which the factor "situation" is applied to the whole plots (each independent set of 15 rsnd generated in one simulation of 2^4 experiment constitutes a whole plot) and each version at each α is applied to one sub-plot. The sub-plots are sets of simulated contrasts, each set being identical

to the set of 15 incremented rsnd produced when the 15 rsnd which constituted the whole plot are modified according to the situation being simulated.

4.3 *Criteria for Evaluating a Version's Performance*

The detection rate D(S) is the average proportion of real contrasts present in situation S which are detected. We use the statistic d(S) to estimate D(S), where

$$d(S) = \sum_{j=0}^{r} j\ p(j)/r, \text{ and}$$

p(j) = (number of simulations in which exactly j of the r real contrasts were detected)/1000.

We need to consider extensions of this criterion to measure a version's detection ability in Type II situations. Obviously, two detection rates, d_1 and d_2, are useful in Type II situations, where

$d_i(r_1,m_1;r_2,m_2)$ = (number of size m_i contrasts detected in the 1000 simulations of $(r_1,m_1;r_2,m_2)$)/1000r_i, $i = 1, 2$.

Another vital aspect of a version's performance is its false positive behavior. One criterion here is the experimentwise error rate (EER). Recall that this is the probability of at least one false positive per experiment in situation S. It is identical to the probability error rate of Miller (1966). We use the statistic fl(S) to estimate the EER, where

$$fl(S) = \sum_{j=1}^{15-r} q(j), \text{ and}$$

$$q(j) = \frac{\text{number of simulations in which exactly } j \text{ of the } 15\text{-}r \text{ null contrasts are declared significant}}{1000} .$$

The sum runs to 15-r because, if r real contrasts are present, it is impossible to declare more than 15-r false positives. Restrictions built into the procedure for a particular half-normal plot version often set the maximum number of false positives at an even smaller number.

Another criterion relating to false positive behavior is the average number of false positives per experiment in situation S, which we refer to as the error rate per experiment (ERPE), using the terminology of Hartley (1955). It is identical to Miller's (1966) expected error rate, if the 15 statements being made about the significance or insignificance of the 15 contrasts in one 2^4 factorial experiment are viewed as a family, in Miller's terminology. We use the statistic f2(S) to estimate the ERPE, where

$$f2(S) = \sum_{j=0}^{15-r} j\, q(j).$$

For evaluating a version's final estimate of σ in a given situation the criteria are the obvious ones: $\bar{s}_f$, the mean of the version's 1000 final estimates of σ in this situation, and $s^2(s_f)$, the variance of these estimates. The first criterion enables us to estimate the bias in the estimates of σ and to observe how this bias changes from situation to situation. The second criterion provides an estimate of the precison of the entire variance estimation process and facilitates efficiency comparisons.

This report concentrates on the measures d, f1, and $\bar{s}_f$. Results for the other measures and two additional versions not reported here are given in Zahn (1969) and are available from the author on request.

5. THE EMPIRICAL BEHAVIORS OF VERSIONS X, S, AND R IN TYPE I SITUATIONS

We discuss the empirical behaviors of versions X, S, and R in the null situation and in 1-situations for α = .05, .20, .40. Since differences among the versions and the performance measures are much the same in 2- and 4-situations as in 1-situations, we consider only one version, version S, in these situations and concentrate on the detection rate. In 6-situations we consider version R, the only one of any use in detecting real contrasts when so many are present.

5.1 *Null Situation Results*

Table 5.1 gives f1, $\bar{s}_f$, and $s(s_f)$ for each version in the null situation at each of three critical value levels: α = .05, .20, and .40. Since the values of f1 are estimates of binomial proportions, the standard deviations of f1 using .05, .20, and .40 level critical values are approximately $\sqrt{.05 \times .95/1000}$ = .007, .013, and .015, respectively. The standard deviation of $\bar{s}_f$ for a particular version and critical value level is easily calculated by dividing the appropriate $s(s_f)$ by $\sqrt{1000}$. Though the differences between several values of f1 and their corresponding α are too large to attribute to chance alone, they are not alarming when we recall that there is also

TABLE 5.1

Empirical Behavior of Versions X, S, and R in the Null Situation Using α = .05, .20, and .40

Version	Criterion	Critical Value Level		
		.05	.20	.40
X	fl	.068	.192	.379
	$\bar{s}_f$	.981	.913	.883
	$s(s_f)$	.193	.218	.238
S	fl	.061	.181	.363
	$\bar{s}_f$	.983	.946	.887
	$s(s_f)$	.192	.214	.236
R	fl	.044	.175	.359
	$\bar{s}_f$	.986	.939	.855
	$s(s_f)$	.196	.230	.273

sampling error in the estimated percentiles being used as critical values. As we expect, σ is underestimated more, on the average, for larger α as more of the larger null contrasts are declared significant and removed from the final estimate of σ. Mildly surprising, however, is the small negative bias (-11.3% to -14.5%) in s_f for all versions using .40 level critical values.

A comparison of $s(s_f)$ to the standard deviation, $1/\sqrt{2n}$, of s, the sample standard deviation, based on n degrees of freedom indicates that the estimate of σ given by the half-normal plot using $\alpha = .05$ is as efficient as an s based on 13.4 "honest" degrees of freedom. By honest, we mean that the variables included in the construction of s are all i.i.d. $N(0,\sigma^2)$. A simulation study not reported here indicated that if in (0) the half-normal plot always uses all 15 contrasts to estimate σ, i.e., if it uses 0.0 level critical values, its final estimate of σ is 99% as efficient as s. Thus, the lower efficiency of the half-normal plot σ estimates using $\alpha = .05$ is not due to the fact that σ is estimated by the slope of the line fitted to the error contrasts, rather than s. Instead, the source of the inefficiency is that each half-normal plot is allowed to declare contrasts significant and remove them from the estimate of σ. Hence, occasionally the half-normal plot's final estimate of σ in (0) will be based on 14 or fewer null contrasts. This is the price we pay for having the power to detect real contrasts and remove them from the final estimate of σ. Using $\alpha = .20$ and $\alpha = .40$ we are obviously more likely to detect real contrasts, but the price we pay in the null situation is that s_f is now only as efficient as an s based on 10.9 and 9.0 degrees of freedom, respectively.

5.2 1-*Situation Results*

Table 5.2 describes the empirical behavior of the versions using $\alpha = .05$, .20, and .40 in 1-situations. We can examine the precision of the results given in Table 5.2 by noting that d and fl are merely estimates of binomial proportions and their variances can be estimated by the usual formulae. The estimated standard error of $\bar{s}_f$, $s(\bar{s}_f)$, is obviously $s(s_f)/\sqrt{1000}$. Specific values of $s(s_f)$ are presented in Zahn (1969). For the versions, critical value levels, and situations in Table 5.2 the values of $s(\bar{s}_f)$ range from a minimum of .006 for all versions using .05 level critical values in situation $(1,1\sigma)$ to a maximum of .011 for version R using .05 level critical values in situation $(1,5\sigma)$, with 90% of the values being in the range .007 to .010.

The differences among the values of d, the detection rate, in situations $(1,1\sigma)$ and $(1,2\sigma)$ for the three versions are small. Versions X and S have considerably larger detection rates than R when the size of the real contrast present is between 3σ and 7σ. Figure 5.1 summarizes the differences in d among the versions for $\alpha = .05$ and .20. It further emphasizes the similarity between versions X and S and the dissimilarity between them and version R.

Through detection rate varies considerably from version to version, all fl values are close to their respective α's. The largest differences between fl and α occur when the real contrast is small. Even in $(1,2\sigma)$ the probability of the real contrast being x_{15} is only 0.45. Thus, the low detection rate in this situation results in few opportunities to declare even one false positive.

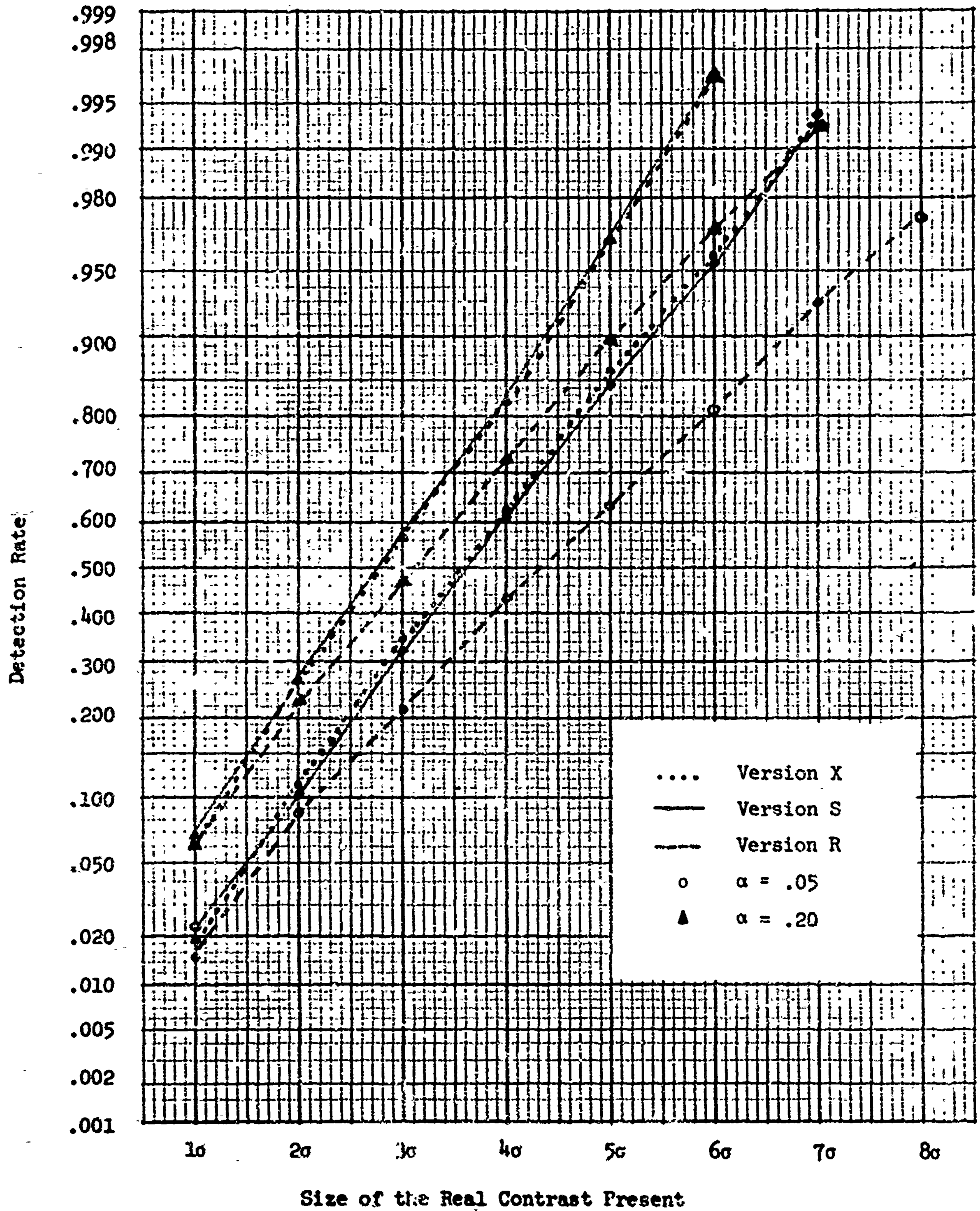
TABLE 5.2

Empirical Behavior of Versions X, S, *and* R *in* 1-*Situations Using* α = .05, .20, *and* .40

Version	Criterion	α	(1,1σ)	(1,2σ)	(1,3σ)	(1,4σ)	(1,5σ)	(1,6σ)	(1,7σ)	(1,8σ)
			Situations							
X	d	.05	.019	.115	.334	.615	.862	.959	.994	1.000
		.20	.060	.266	.564	.820	.966	.997	1.000	1.000
		.40	.121	.393	.730	.923	.991	.999	1.000	1.000
	fl	.05	.047	.040	.040	.047	.043	.051	.047	.042
		.20	.153	.165	.153	.192	.174	.176	.171	.191
		.40	.340	.349	.346	.359	.377	.346	.340	.364
	$\bar{s}_f$	.05	1.024	1.059	1.098	1.084	1.031	.999	.985	.993
		.20	.985	.990	1.006	.978	.954	.949	.948	.951
		.40	.923	.914	.920	.900	.891	.898	.890	.895
S	d	.05	.021	.105	.310	.605	.849	.957	.994	1.000
		.20	.063	.264	.566	.826	.965	.998	1.000	1.000
		.40	.117	.399	.714	.911	.989	.999	1.000	1.000
	fl	.05	.046	.040	.043	.057	.049	.049	.047	.050
		.20	.156	.166	.154	.207	.182	.198	.161	.196
		.40	.317	.358	.351	.388	.383	.367	.354	.384
	$\bar{s}_f$	.05	1.024	1.062	1.104	1.086	1.034	1.001	.985	.990
		.20	.981	.990	1.004	.972	.952	.942	.949	.946
		.40	.928	.915	.920	.898	.887	.891	.896	.889
R	d	.05	.015	.089	.216	.436	.638	.809	.928	.974
		.20	.064	.230	.469	.728	.897	.971	.993	1.000
		.40	.143	.412	.564	.877	.969	.998	1.000	1.000
	fl	.05	.038	.046	.043	.038	.048	.034	.040	.048
		.20	.160	.183	.149	.193	.187	.179	.160	.200
		.40	.310	.366	.357	.401	.372	.389	.350	.403
	$\bar{s}_f$	.05	1.026	1.067	1.137	1.167	1.139	1.093	1.031	1.011
		.20	.972	.989	1.025	1.005	.962	.946	.942	.931
		.40	.900	.881	.902	.875	.862	.849	.861	.840

FIGURE 5.1

Detection Rates for Versions X, S, *and* R *in*
1-Situations Using α = .05 *and* .20



All versions have a tendency to overestimate σ, particularly when small real contrasts are present. The bias is worst in $(1,4\sigma)$ and, for the better versions, is negligible in $(1,6\sigma)$ and $(1,8\sigma)$ for $\alpha = .05$. A striking aspect of $\bar{s}_f$'s behavior for all versions is that $\bar{s}_f$ is not a monotone function of the size of the real contrast. Rather, $\bar{s}_f$ increases as the size of the real contrast increases to 3σ or 4σ, depending on the version, and then decreases as the size increases to 8σ. To explain this, we note that while there are fewer undetected contrasts in $(1,4\sigma)$ than in $(1,2\sigma)$, the bias caused by an undetected contrast is greater if its size is 4σ than if its size is 2σ. The additional bias offsets the fact that the detection rate is greater in $(1,4\sigma)$ than in $(1,2\sigma)$.

Version R's inferior detection rate affects its final estimate of σ in two rather obvious ways: (1) Any undetected real contrast will be included in construction of s_f. Since R detected the fewest real contrasts, the bias caused by undetected real contrasts will be more severe for R than for the other versions. (2) The estimates of σ vary more for version R than they do for the other versions; thus, $s(s_f)$ is larger for R than for the other versions.

5.3 2- *and* 4-*Situation Results*

Since the differences among the versions and criteria are much the same in 2- and 4-situations as in 1-situations, we concentrate here on the detection rate, perhaps the criterion of most interest to the experimenter, and version S. We concentrate on S since it is generally superior to the other versions, especially in 4-situations where its detection rate always exceeded the detection rates of

the other versions, often by as much as .07 to .15. Empirical results for other versions and other criteria in these and other situations appear in Zahn (1969) and are available from the author on request.

Table 5.3 gives empirical estimates of the detection rate, EER, and average final estimate of σ for version S in 1-, 2-, and 4-situations using α = .05, .20, and .40. Table 5.3 also gives the estimated standard errors of d and $\bar{s}_r$. To compute the standard error of d in 2-, 4-, and 6-situations, we first note that results for individual real contrasts are not independent within trials. Thus, d behaves as a proportion estimated by cluster sampling and its standard error can be estimated using the appropriate formulae in Cochran (1963, p. 64).

As we expect, for a fixed number of real contrasts present, the detection rate increases as the size of the real contrasts increases. However, the detection rate decreases as the number of real contrasts present increases. Version S has a moderately smaller detection rate in 2-situations than in 1-situations. It has a much smaller detection rate in 4-situations than in 1-situations.

Examining Table 5.3 we see that increasing α from .05 to .20 yields a sizable increase in the detection rate of version S. Increasing α to .40 yields even larger detection rates. The price we pay for the larger detection rates is, of course, that the probability of at least one false positive is much larger. However, another benefit helping to offset this cost is that the bias in $\bar{s}_r$ decrease sharply as α increases.

TABLE 5.3

Empirical Behavior of Version S in 1-, 2-, and 4-Situations with Real Contrasts of Sizes 2σ, 4σ, 6σ, and 8σ Present, Using α = .05, .20, and .40.

Number of Real Contrasts Present	Criterion	α	Size of the Real Contrasts Present			
			2σ	4σ	6σ	8σ
1	d, s.e.(d)	.05	.105, .010	.605, .015	.957, .006	1.000, .000
		.20	.264, .014	.826, .012	.998, .001	1.000, .000
		.40	.399, .015	.911, .009	.999, .001	1.000, .000
	f1	.05	.040	.057	.049	.050
		.20	.166	.207	.198	.196
		.40	.358	.386	.367	.384
	$\bar{s}_f$, s.e.($\bar{s}_f$)	.05	1.062, .007	1.086, .010	1.001, .008	.990, .007
		.20	.990, .008	.972, .009	.942, .007	.946, .007
		.40	.915, .008	.898, .008	.891, .008	.889, .008
2	d, s.e.(d)	.05	.080, .007	.539, .014	.954, .006	1.000, .000
		.20	.219, .011	.821, .010	.997, .001	1.000, .000
		.40	.365, .012	.920, .007	1.000, .001	1.000, .000
	f1	.05	.023	.044	.046	.058
		.20	.121	.166	.181	.188
		.40	.283	.341	.335	.358
	$\bar{s}_f$, s.e.($\bar{s}_f$)	.05	1.171, .008	1.233, .013	1.019, .010	.976, .007
		.20	1.078, .009	1.016, .010	.942, .007	.939, .007
		.40	.984, .009	.921, .008	.901, .007	.893, .007
4	d, s.e.(d)	.05	.037, .004	.270, .01[illegible]	.864, .010	.997, .002
		.20	.134, .007	.654, .013	.988, .003	1.000, .000
		.40	.247, .009	.835, .009	.999, .001	1.000, .000
	f1	.05	.008	.002	.000	.000
		.20	.050	.010	.000	.000
		.40	.143	.028	.000	.000
	$\bar{s}_f$, s.e.($\bar{s}_f$)	.05	1.389, .008	1.846, .020	1.242, .024	1.001, .008
		.20	1.282, .010	1.326, .016	1.011, .009	.994, .007
		.40	1.169, .001	1.121, .013	0.995, .007	.994, .007

5.4 6-*Situation Results*

Table 5.4 summarizes the performance of version R using $\alpha = .05$, .20, and .40 in 6-situations. Only version R does not break down in 6-situations. When six real contrasts are present, the test statistic denominators of versions X and S will either equal or include the smaller real contrasts. This contamination always occurs and is severe enough so that neither version detects even 2% of the real contrasts present. Since the test statistic denominators of version R are based on at most the seven smallest order statistics, R still detects some real contrasts in 6-situations. Of course, its detection rate is smaller in 6-situations than in 4-situations.

The gap between α and fl is wider in 6-situations than in 1-, 2-, or 4-situations. In addition, this gap narrows as the size of the real contrasts present increases.

The bias in $\bar{s}_r$ which is severe in 6-situations is smallest for version R since it has the largest detection rate in these situations. However, even for version R the bias is large. Furthermore, s_r for R in 6-situations is exceedingly variable, which is not surprising since s_r will equal approximately 5.0 if none of the real contrasts in $(6,8\sigma)$ are detected and approximately 1.0 if all the real contrasts in this situation are detected.

In these situations the bias in $\bar{s}_r$ can be greatly reduced and the detection rate dramatically increased by using larger α. Hence, we highly recommend the half-normal plot with $\alpha = .20$ or .40 level critical values to the experimenter who is doing exploratory research and might encounter a 4- or a 6-situation.

TABLE 5.4

Empirical Behavior of Version R in 6-Situations

Using α = .05, .20, *and* .40

Version	Criterion	α	Situations (6,2σ)	(6,4σ)	(6,6σ)	(6,8σ)
		.05	.015, .003	.132, .010	.430, .015	.761, .013
	d,s.e.(d)	.20	.074, .006	.412, .014	.837, .001	.983, .004
		.40	.189, .009	.597, .013	.976, .004	1.000, .000
		.05	.005	.020	.025	.047
R	fl	.20	.041	.129	.190	.177
		.40	.135	.306	.378	.385
		.05	1.603, .009	2.454, .021	2.602, .047	1.938, .055
	$\acute{s}_f$,s.e.($\bar{s}_f$)	.20	1.516, .011	1.889, .029	1.340, .035	1.024, .018
		.40	1.360, .014	1.355, .026	.979, .016	.913, .008

5.5. *On Daniel's Questions about Plot Modifications*

The results of our research suggest answers to some questions raised by Daniel (1959) on possible variants of the half-normal plot. He wonders if "an invariable rule should be set up, using only some fixed proportion of the smaller contrasts to estimate error". We oppose the idea of using a fixed number of contrasts in a single replicate 2^4 factorial experiment to estimate error because of its inefficiency when there are only one or two real contrasts.

Another query is whether one should use the error contrasts to form a mean square error term. Since fitting a line to the error contrasts yields a highly efficient, quick-and-easy estimate of σ, we do not feel that it is necessary to form the mean square error term.

Daniel also questions if one should "decline to use only higher-order interactions for error since some plot-splitting is almost inevitable in multi-stage processes". This seems wise if the danger of hidden plot-splitting is sizable, though this analysis would present many other complications as well.

We feel that one should "insist on at least partial duplication of 2^{p-q} experiments when no good previous estimate of error is available" (Daniel, 1959), especially when the experimenter thinks that as many as four real contrasts may be present when $p-q = 4$. Without the partial duplication in these difficult situations, the error variation estimate is badly biased when several real contrasts of any size, or a few small-to-medium-sized real contrasts, are present.

An alternative procedure which has been suggested for use when as many as many as six or nine real contrasts are present in a non-replicated 2^4 factorial experiment is the chain-pooling ANOVA (Holms and Berrettoni, 1969). In these situations the chain pooling procedure might be superior to the half-normal plot versions discussed in this paper.

5.6 *The Half-Normal Plot as an Outlier Rejection Procedure*

Suppose we observe 15 random variables which are thought to be i.i.d. $N(0,\sigma^2)$ in order to estimate σ^2. However, if some of the observations are outliers with means $\mu \neq 0$ we will want to exclude these observations from the estimation of σ^2. Now, the location of outliers under these circumstances poses the same problem as does the detection of real contrasts in a single replicate 2^4 factorial experiment. Thus, the half-normal plot can also be used as an outlier rejection procedure.

While doing pilot studies for the Main Study, we examined the power of the outlier rejection procedure (BCT) proposed by Bliss, Cochran, and Tukey (1956). The pilot study results demonstrated a serious defect in this procedure. Although the BCT procedure is reasonably sensitive to outliers when only two outliers are present, it is almost useless as an outlier rejection procedure when three or four are present. For example, when four outliers, each distributed $N(6\sigma,\sigma^2)$, are present in the situation described in the previous paragraph, BCT detects approximately 12% of them. Since all fifteen observations are used in the denominator of BCT's rejection criteria, outliers will always contaminate the denominator. The consequences of this contamination are most serious.

Obviously, BCT is more adversely affected by an increase in the number of outliers than is the half-normal plot. A suggestion based on the half-normal plot results is to modify the BCT denominator so that it does not include the larger observations. For instance, in the situation described at the beginning of this section, a modified denominator which might be of interest is the sum of the eleven observations closest to 0.0. This modification should make the procedure more robust, though it will sacrifice some efficiency when only one outlier is present.

We feel that there are inadequate warnings in the statistical literature as to the dire consequences such as the above which may result from including outliers in the denominators of the test statistic. Several of the conventional outlier rejection procedures include all observations in the test statistic denominators. For instance, when searching for one outlier, Grubbs (1969) recommends $T_n = \frac{x_n - \bar{x}}{s}$, where x_n = the largest observation in the sample,

$$\bar{x} = \sum_{1}^{n} x_i / n, \qquad s = \sqrt{\frac{\sum_{1}^{n}(x_i - \bar{x})^2}{n-1}} \quad .$$

We question whether many experimenters appreciate how drastically the sensitivity of procedures such as T_n may be affected by the inflation of the test statistic denominator which occurs when two or three outliers are included in it. In general, our suggestion is to base the test statistic denominator on only the smaller observations in order to minimize the probability of contaminating the denominator.

6. THE EMPIRICAL BEHAVIORS OF VERSIONS X, S, AND R IN TYPE II SITUATIONS

This section discusses the results of using the half-normal plot to analyze experiments in Type II situations, i.e., situations in which there are real contrasts of two different sizes. The main results for Type II situations are summarized in Tables 6.1 and 6.2.

6.1 1-1-*Situation Results*

In order to isolate the effect of the presence of one size m_2 contrast on the detection rate for one size m_1 contrast in situation $(1,m_1;1;m_2)$, we have constructed Table 6.1. Consider the section of this table devoted to version X. The five rows of this section represent detection rate curves for version X under five different sets of conditions. The first row gives $d(1,m)$, for $m = 2\sigma$, 4σ, 6σ, and 8σ in 1-situations; the second row gives $d_1(1,m_1;1;2\sigma)$ for $m_1 = 2\sigma, 4\sigma, 6\sigma, 8\sigma$; etc.

To understand how the detection rate behaves in 1-1-situations, we examine how $d_1(1,4\sigma;1,m_2)$ varies as m_2 varies from 0σ to 8σ by considering the second column of Table 6.1. This shows that a size 4σ contrast is more likely to be detected if it is the only real contrast present than if another real contrast is present. This is reasonable since the detection rate of all versions has been observed to decline as the number of real contrasts present increases. As a second real contrast of increasing size is introduced, we note that the detection rate for a size 4σ contrast drops at first from 0.44 to 0.35 for R and then rises to 0.41 as the size of the second contrast

TABLE 6.1

Detection Rate for One Size m_1 Contrast When One Size m_2 Contrast is Also Present in the Experiment

These results are all for versions using 0.05 level critical values.

Version	m_2 \ m_1 =	2σ	4σ	6σ	8σ
	0σ	0.09[1]	0.44[1]	0.81[1]	0.97[1]
	2σ	0.06[2]	0.35	0.75	0.95
R	4σ	0.07	0.38[2]	0.76	0.96
	6σ	0.08	0.41	0.81[2]	0.96
	8σ	0.07	0.41	0.81	0.96[2]
	0σ	0.12[1]	0.62[1]	0.96[1]	1.00[1]
	2σ	0.09[2]	0.51	0.93	1.00
X	4σ	0.10*	0.53[2]	0.93	0.99
	6σ	0.09	0.57	0.95[2]	1.00
	8σ	0.10	0.6[illegible]	0.96	1.00[2]
	0σ	0.11[1]	0.62[1]	0.96[1]	1.00[1]
S	2σ	0.08[2]	0.51	0.92	1.00
	4σ	0.10	1.54[2]	0.94	1.00
	6σ	0.10	0.58	0.95[2]	1.00
	8σ	0.10	0.62	0.96	1.00[2]

[1] These detection rates are the detection rates in the respective 1-situations.

[2] These detection rates are the detection rates in respective 2-situations.

* $0.10 = d_1(1,2\sigma;1,4\sigma)$ = Detection rate for the size 2σ contrast in situation $(1,2\sigma;1,4\sigma)$, i.e., 10% of the size 2σ contrasts present in the 1000 simulations of situation $(1,2\sigma;1,4\sigma)$ were detected.

increases to 8σ. The explanation of the rise is that as m_2 increases in size, the second real contrast is more likely to be detected. Consequently, the size 4σ contrast is examined for significance more often and is more likely to be detected. The detection rate for the size 4σ contrast in the presence of an additional, large real contrast is, however, less than the detection rate in $(1,4\sigma)$. The reason is as follows. In $(1,4\sigma)$ the real contrast is present with 14 null contrasts, whereas here the size 4σ contrast is present with one large real contrast, i.e., the second real contrast, and only 13 null contrasts. Thus, less information to estimate σ is available than in $(1,4\sigma)$ and we expect to see a slightly smaller detection rate for the size 4σ contrast than in $(1,4\sigma)$, even when the second real contrast is large. The dips and subsequent rises in detection rate occur, to within sampling errors, for every version and every size contrast. Though consistent, these dips and rises are not large.

Similar comments hold for the 2-2-situation results which are given in Table 6.2. However, in these situations the dips and rises previously noted are large.

7. NOMINATION AND THE HALF-NORMAL PLOT: SOME COMPARISONS

In analyzing experiments lacking a classical, internal estimate of error variance, another approach is to decide a priori to combine the higher order interactions to form an estimate of error variance. This approach, which we refer to as "nomination", has been widely used by experimenters doing single replicate factorial experiments or

TABLE 6.2

Detection Rate for the Two Size m_1 Contrasts When Two Size m_2 Contrasts are Also Present in the Experiment

These results are all for versions using 0.05 level critical values.

Version	m_2 \ m_1 =	2σ	4σ	6σ	8σ
R	0σ	0.06[2]	0.38[2]	0.81[2]	0.96[2]
	2σ	0.04[4]	0.24	0.63[*]	0.90
	4σ	0.04	0.25[4]	0.60	0.88
	6σ	0.06	0.31	0.67[4]	0.89
	8σ	0.05	0.32	0.72	0.91[4]
X	0σ	0.09[2]	0.53[2]	0.95[2]	1.00[2]
	2σ	0.04[4]	0.29	0.82	0.99
	4σ	0.05	0.21[4]	0.61	0.91
	6σ	0.08	0.40	0.72[4]	0.91
	8σ	0.07	0.45	0.86	0.95[4]
S	0σ	0.08[2]	0.54[2]	0.95[2]	1.00[2]
	2σ	0.04[4]	0.33	0.85	1.00
	4σ	0.06	0.27[4]	0.77	0.98
	6σ	0.08	0.49	0.86[4]	0.99
	8σ	0.07	0.54	0.95	1.00[4]

[2] These detection rates are the detection rates in the respective 2-situations.

[4] These detection rates are the detection rates in the respective 4-situations.

[*] 0.63=d_2(2,2σ;2,6σ) = Detection rate for the size 6σ contrasts in situation (2,2σ;2,6σ), i.e., 81% of the size 6σ real contrasts present in the 1000 simulations of situation (2,2σ;2,6σ) were detected.

their fractions in well-researched areas, such as agriculture, where there is ample evidence from earlier experiments that the real effects of such high order interactions are usually negligible.

A nomination procedure which is illustrated by an example in Davies (1954, p. 274) consists of the following steps:

(1) The experimenter assumes that certain contrasts, the "nominated" contrasts, are null and constructs s_N^2, an estimate of σ^2, from them.

(2) Each of the remaining contrasts is tested for significance by dividing its square by s_N^2 and comparing the result to a percentage point of the F-distribution.

Since the results of the Main Study indicate that, barring the breakdown of a procedure, increasing the EER and ERPE results in an increase in detection rate, we shall compare in this section the half-normal plot and a nomination procedure with similar EER and ERPE. For the EER of the nomination procedure to be comparable to the EER of the half-normal plot using .05 level critical values, the 0.5% percentage point of the appropriate F should be used, while the 1.0% point should be used if we desire the ERPE's of the two procedures to be comparable.

Another difficulty arises while attempting to compare nomination to the half-normal plot: In order to calculate the detection rate of the nomination procedure from tables of the noncentral F- or t-distribution, we must make the basic assumption that the experimenter nominated only null contrasts. This assumption biases the results in

favor of nomination. If it is true, all the real contrasts will be tested for significance. This assumption precludes any contamination by real contrasts of either the test statistic denominator used by nomination or the final estimate of error variance given by nomination. Furthermore, s_N^2 will not become increasingly inflated as the number of real contrasts increases.

We let $N(e,F(1,e,p))$ denote the variant of nomination in which e contrasts are nominated and $F(1,e,p)$ is used as the critical value.

In this section we restrict our attention to 2^4 factorial experiments and two nomination procedures: one nominating five error contrasts, $N(5,F)$, and the other nominating ten, $N(10,F)$. By interpolation in Tang's tables (1938) and in the non-central t-tables of Resnikoff and Lieberman (1957), we can calculate the power of $N(5,F)$ and $N(10,F)$ in $(1,2\sigma)$, $(1,4\sigma)$, $(1,6\sigma)$ using various F-percentage points as critical values. These results are given in Table 7.1.

Although the half-normal plot has been unfavorably compared to nomination procedures in this section, it gives a very good account of itself with respect to detection rate when compared to nomination procedures with similar EER and ERPE. The results of this section indicate that the half-normal plot has a distinctly larger detection rate than a nomination procedure using the same EEP if the experimenter's prior information will only allow him to nominate 3- and 4-factor interactions. However, if he can accurately nominate ten error contrasts, nomination will have a larger detection rate than the half-normal plot if four real contrasts are present.

TABLE 7.1

Detection Rate of N(5,F) and N(10,F) Using Three Critical Values

N(5,F(1,5,·))

(Nominating all 3- and 4-factor interactions)

Critical Value	(1,2σ)	Situation (1,4σ)	(1,6σ)
F(1,5,.05)	.34	.86	.99
F(1,5,.01)	.13	.55	.91
F(1,5,.005)	.08	.39	.79

N(10,F(1,10,.))

(Nominating all 3- and 4-factor interactions and 5 of the 6 2-factor interactions)

Critical Value	(1,2σ)	Situation (1,4σ)	(1,6σ)
F(1,10,.05)	.44	.93	1.00
F(1,10,.01)	.19	.77	.99
F(1,10,.005)	not tabled	.65	.97

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 *The Half-Normal Plot and 2^4 Factorial Experiments*

Since the half-normal plot is intended to indicate which contrasts are real and to estimate σ, we will judge it by these standards. As regards detection rate, we observe in Section 5 that the half-normal plot using .05 level critical values has a detection rate as large as 0.12, 0.62, 0.96, and 1.00 in 1-situations for contrasts of size 2σ, 4σ, 6σ, and 8σ, respectively; 0.09, 0.54, 0.95, and 1.00 in 2-situations; 0.04, 0.30, 0.86, and 1.00 4-situations; and 0.01, 0.13, 0.43, and 0.76 in 6-situations. Here, we are reporting only the results for the version having the highest detection rate in each situation. These results lead to our conclusion that the half-normal plot is a suitable procedure for analyzing 2^4 factorial experiments, provided that four or fewer real contrasts are present.

The decline in detection rate as the number of real contrasts increases should be noted. The most drastic decrease in detection rate occurs as the number of real contrasts increases to six. In 6-situations the only version with any detection rate at all is R; its detection rate is reported in the previous paragraph. The other versions have a detection rate of at most .03 in 6-situations.

8.2 *A comparison of the Half-Normal Plot Versions in Various Situations*

In situation (0) all versions are quite similar.

In 1-situations versions X and S are similar to each other and superior to version R in every way: they have larger detection rates and yield less biased, less variable, final estimates of σ.

In 2-situations version R again has little to recommend it. However, it does compare to the other versions slightly better in 2-situations than in 1-situations. The performances of versions X and S are similar.

In 4-situations version S is the best version in terms of detection rate and estimation of σ.

In 6-situations there is little to recommend any version. Version R is the only one which does not collapse. However, its detection rate is much smaller than it was in 4-situations and its final estimate of σ is badly biased.

Of the three half-normal plot versions considered we recommend version S on the basis of its steady performance in 1-, 2-, and 4-situations. If the experimenter expects more than four real contrasts, we advise him to consider whether he can afford an EER of .20 or .40. If he can, we recommend version R with .20 or .40 level critical values. Before acting on this recommendation, the experimenter would be well advised to consider whether a second replicate or a larger fractional replicate might pay for itself by dramatically increasing the detection rate.

8.3 *Nomination versus the Half-Normal Plot*

As described in Section 7, nomination has a smaller detection rate than versions of the half-normal plot with equivalent EER's, unless the experimenter can accurately nominate ten error contrasts. The half-normal plot estimates σ more efficiently than N(5,F) when the real contrasts are large (8σ). However, if the contrasts are

only medium-sized and if the nomination is accurate, N(5,F) is more efficient than the half-normal plot versions examined in the Main Study. The procedure N(10,F) is as efficient as the half-normal plot even when the real contrasts are large.

If equivalent EER's are desired, our recommendation is to use the half-normal plot unless almost all null contrasts can be accurately nominated.

ACKNOWLEDGMENTS

Research for this article was conducted while the author was a graduate student in the Department of Statistics, Harvard University, and was supported by the Office of Naval Research Contract N00014-67-A-0298-0017 and the National Science Foundation. During revisions at Florida State University the author was partially supported by Biometry Training Grant 5 T01 GM913 from the National Institute of General Medical Sciences.

REFERENCES

Birnbaum, A. (1959). On the analysis of factorial experiments with replication. Technometrics, 1, 342-357.

Blankenskip, J. (1965). Use of order statistics for graphical estimation. Ph.D. Thesis, Oklahoma State University.

Bliss, C. I., Cochran, W. G., and Tukey, J. W. (1956). A rejection criterion based on the range. Biometrika, 43, 418-422.

Cochran, W. G. (1963). Sampling Techniques. Second Edition, John Wiley, New York.

Daniel, C. (1959). Use of half-normal plots in interpreting factorial two-level experiments. Technometrics, 1, 311-341.

Davies, O. L. (1954) (Editor). Design and Analysis of Industrial Experiments. Second Edition, Oliver and Boyd, London, and Hafner Publishing Company, New York.

Ferguson, T. S. (1960). Discussion of the papers of Messrs. Anscombe and Daniel. Technometrics, 2, 159.

Hartley, H. O. (1955). Some recent developments in analysis of variance. Comm. Pure Appl. Math., 8, 47-72.

Holms, A. G., and Berrettoni, J. N. (1969). Chain-pooling ANOVA for two-level factorial replication-free experiments. Technometrics, 11, 725-746.

Miller, R. G. (1966). Simultaneous Statistical Inference. McGraw-Hill Book Company, New York.

Resnikoff, G. J., and Lieberman, G. J. (1957). Tables of the Non-central t-Distribution. Stanford University Press, Stanford, California.

Tang, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. Statist. Res. Mem., 2, 126-149.

Wilks, S. S. (1962). Mathematical Statistics. John Wiley, New York.

Zahn, D. A. (1969). An empirical study of the half-normal plot. Ph.D. Thesis, Harvard University.